HELDA
HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

UNIVERSITY OF HELSINKI

https://helda.helsinki.fi

A large-scale genome-wide association study meta-analysis of cannabis use disorder

the Psychiatric Genomics Consortium Substance Use Disorders Workgroup

2020

the Psychiatric Genomics Consortium Substance Use Disorders Workgroup , Johnson , E C , Demontis , D & Thorgeirsson , T E 2020 , ' A large-scale genome-wide association study meta-analysis of cannabis use disorder ' , The Lancet psychiatry , vol. 7 , no. 12 , pp. 1032-1045 . https://doi.org/10.1016/S2215-0366(20)30339-4

http://hdl.handle.net/10138/326950
https://doi.org/10.1016/S2215-0366(20)30339-4

cc_by_nc_nd
publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

A large-scale genome-wide association study meta-analysis of cannabis use disorder

Emma C Johnson, Ditte Demontis*, Thorgeir E Thorgeirsson*, Raymond K Walters, Renato Polimanti, Alexander S Hatoum, Sandra Sanchez-Roige, Sarah E Paul, Frank R Wendt, Toni-Kim Clarke, Dongbing Lai, Gunnar W Reginsson, Hang Zhou, June He, David A A Baranger, Daniel F Gudbjartsson, Robbee Wedow, Daniel E Adkins, Amy E Adkins, Jeffry Alexander, Silviu-Alin Bacanu, Tim B Bigdeli, Joseph Boden, Sandra A Brown, Kathleen K Bucholz, Jonas Bybjerg-Grauholm, Robin P Corley, Louisa Degenhardt, Danielle M Dick, Benjamin W Domingue, Louis Fox, Alison M Goate, Scott D Gordon, Laura M Hack, Dana B Hancock, Sarah M Hartz, Ian B Hickie, David M Hougaard, Kenneth Krauter, Penelope A Lind, Jeanette N McClintick, Matthew B McQueen, Jacquelyn L Meyers, Grant W Montgomery, Ole Mors, Preben B Mortensen, Merete Nordentoft, John F Pearson, Roseann E Peterson, Maureen D Reynolds, John P Rice, Valgerdur Runarsdottir, Nancy L Saccone, Richard Sherva, Judy L Silberg, Ralph E Tarter, Thorarinn Tyrfingsson, Tamara L Wall, Bradley T Webb, Thomas Werge, Leah Wetherill, Margaret J Wright, Stephanie Zellers, Mark J Adams, Laura J Bierut, Jason D Boardman, William E Copeland, Lindsay A Farrer, Tatiana M Foroud, Nathan A Gillespie, Richard A Grucza, Kathleen Mullan Harris, Andrew C Heath, Victor Hesselbrock, John K Hewitt, Christian J Hopfer, John Horwood, William G Iacono, Eric O Johnson, Kenneth S Kendler, Martin A Kennedy, Henry R Kranzler, Pamela A F Madden, Hermine H Maes, Brion S Maher, Nicholas G Martin, Matthew McGue, Andrew M McIntosh, Sarah E Medland, Elliot C Nelson, Bernice Porjesz, Brien P Riley, Michael C Stallings, Michael M Vanyukov, Scott Vrieze, the Psychiatric Genomics Consortium Substance Use Disorders Workgroup, Lea K Davis, Ryan Bogdan, Joel Gelernter†, Howard J Edenberg†, Kari Stefansson†, Anders D Børglum†, Arpana Agrawal†*

Lancet Psychiatry 2020; 7: 1032–45

Published Online October 20, 2020 https://doi.org/10.1016/S2215-0366(20)30339-4

See Comment page 1002

*Contributed equally

†Jointly supervised this work

Department of Psychiatry (E C Johnson PhD, A S Hatoum PhD, J He MSc, Prof K K Bucholz PhD, L Fox BS, S M Hartz MD, Prof J P Rice PhD, Prof L J Bierut MD, Prof R A Grucza PhD, Prof A C Heath DPhil, Prof P A F Madden PhD, Prof E C Nelson MD, Prof A Agrawal PhD) **and Department of Genetics** (N L Saccone PhD), **Division of Biostatistics** (N L Saccone), **Washington University School of Medicine, St Louis, MO, USA; Department of Biomedicine—Human Genetics and Centre for Integrative Sequencing** (D Demontis PhD, Prof A D Børglum MD), **and National Centre for Register-Based Research** (Prof P B Mortensen MD), **Centre for Integrated Register-based Research** (Prof P B Mortensen), **Aarhus University, Aarhus, Denmark; The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Aarhus, Denmark** (D Demontis, J Bybjerg-Grauholm MSc, D M Hougaard MD,

Summary

Background Variation in liability to cannabis use disorder has a strong genetic component (estimated twin and family heritability about 50–70%) and is associated with negative outcomes, including increased risk of psychopathology. The aim of the study was to conduct a large genome-wide association study (GWAS) to identify novel genetic variants associated with cannabis use disorder.

Methods To conduct this GWAS meta-analysis of cannabis use disorder and identify associations with genetic loci, we used samples from the Psychiatric Genomics Consortium Substance Use Disorders working group, iPSYCH, and deCODE (20 916 case samples, 363 116 control samples in total), contrasting cannabis use disorder cases with controls. To examine the genetic overlap between cannabis use disorder and 22 traits of interest (chosen because of previously published phenotypic correlations [eg, psychiatric disorders] or hypothesised associations [eg, chronotype] with cannabis use disorder), we used linkage disequilibrium score regression to calculate genetic correlations.

Findings We identified two genome-wide significant loci: a novel chromosome 7 locus (*FOXP2*, lead single-nucleotide polymorphism [SNP] rs7783012; odds ratio [OR] 1·11, 95% CI 1·07–1·15, $p=1·84\times10^{-9}$) and the previously identified chromosome 8 locus (near *CHRNA2* and *EPHX2*, lead SNP rs4732724; OR 0·89, 95% CI 0·86–0·93, $p=6·46\times10^{-9}$). Cannabis use disorder and cannabis use were genetically correlated (r_g 0·50, $p=1·50\times10^{-21}$), but they showed significantly different genetic correlations with 12 of the 22 traits we tested, suggesting at least partially different genetic underpinnings of cannabis use and cannabis use disorder. Cannabis use disorder was positively genetically correlated with other psychopathology, including ADHD, major depression, and schizophrenia.

Interpretation These findings support the theory that cannabis use disorder has shared genetic liability with other psychopathology, and there is a distinction between genetic liability to cannabis use and cannabis use disorder.

Funding National Institute of Mental Health; National Institute on Alcohol Abuse and Alcoholism; National Institute on Drug Abuse; Center for Genomics and Personalized Medicine and the Centre for Integrative Sequencing; The European Commission, Horizon 2020; National Institute of Child Health and Human Development; Health Research Council of New Zealand; National Institute on Aging; Wellcome Trust Case Control Consortium; UK Research and Innovation Medical Research Council (UKRI MRC); The Brain & Behavior Research Foundation; National Institute on Deafness and Other Communication Disorders; Substance Abuse and Mental Health Services Administration (SAMHSA); National Institute of Biomedical Imaging and Bioengineering; National Health and Medical Research Council (NHMRC) Australia; Tobacco-Related Disease Research Program of the University of California; Families for Borderline Personality Disorder Research (Beth and Rob Elliott) 2018 NARSAD Young Investigator Grant; The National Child Health Research Foundation (Cure Kids); The Canterbury Medical Research Foundation; The New Zealand Lottery Grants Board; The University of Otago; The Carney Centre for Pharmacogenomics; The James Hume Bequest Fund; National Institutes of Health: Genes, Environment and Health Initiative; National Institutes of Health; National Cancer Institute; The William T Grant Foundation; Australian

Research Council; The Virginia Tobacco Settlement Foundation; The VISN 1 and VISN 4 Mental Illness Research, Education, and Clinical Centers of the US Department of Veterans Affairs; The 5th Framework Programme (FP-5) GenomEUtwin Project; The Lundbeck Foundation; NIH-funded Shared Instrumentation Grant S10RR025141; Clinical Translational Sciences Award grants; National Institute of Neurological Disorders and Stroke; National Heart, Lung, and Blood Institute; National Institute of General Medical Sciences.

Copyright © 2020 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license.

Introduction

Cannabis use is common, but most users do not progress to cannabis use disorders. About 50–70% of liability to cannabis use disorders is due to genetic factors.[1] Three genome-wide association studies (GWASs) of cannabis use disorders[2–4] have identified variants reaching genome-wide significance, but inadequate sample sizes (sample size from largest study to date: 51 372, with 2387 cases) and heterogeneity among samples have contributed to a paucity of replicable findings: only one locus, tagged by a *cis*-eQTL for *CHRNA2* (encoding a nicotinic acetylcholine receptor), has been robustly identified.[3]

A GWAS of lifetime cannabis use (184 765 total sample size, 43 380 cases) identified eight genome-wide significant loci and 35 significant genes.[5] Twin studies suggest high genetic correlations between early stages of cannabis experimentation and later cannabis use disorder.[6] However, casual cannabis use is affected by a variety of socioenvironmental influences and age-period-cohort effects, whereas progression to cannabis use disorder is related to other psychopathologies. Findings have suggested partially distinct genetic causes underlying alcohol consumption and alcohol use disorder, including different genetic associations with other psychiatric disorders and traits.[7,8] Thus, in addition to examining the genomic liability for cannabis use disorder, we tested whether the genetic influences underlying cannabis use and cannabis use disorder diverge with respect to behavioural and brain measures.

Research in context

Evidence before this study

Cannabis use disorder is heritable (50–70% according to twin and family studies), yet identification of genomic variants associated with cannabis use disorder from genome-wide association studies (GWASs) remains sparse. We surveyed all peer-reviewed journal publications in English on GWASs of cannabis use disorder or cannabis dependence using Google Scholar and PubMed, published between Jan 1, 1990, and April 1, 2020. Search terms included "cannabis dependence", "cannabis abuse", "cannabis use disorder", "marijuana dependence", "marijuana abuse", "marijuana use disorder", and "GWAS". The most promising finding to date is a variant that is a *cis*-eQTL for *CHRNA2* (Demontis and colleagues), which was replicated in an independent dataset for cannabis use disorder. Independently, GWAS of cannabis use have identified multiple genetic risk loci; however, the extent to which the genetics of cannabis use correlates with liability to cannabis use disorder has not been determined. Although GWASs of cannabis use have been studied in the context of a variety of psychiatric and psychosocial correlates, it is expected that some divergent associations will be seen when looking at cannabis use disorder. Previous studies have drawn causal links between cannabis exposure and brain volume, but the relationship between genetic liability to cannabis use disorder and brain volume in individuals naive to cannabis has not yet been studied.

Added value of this study

Our study is the current largest GWAS of cannabis use disorder and the first to include a transancestral component. We found a novel risk locus on chromosome 7. The lead risk variant at this locus is an eQTL for *FOXP2*—a gene previously implicated in risk-taking behaviours. Contrasting cannabis use and cannabis use disorder, we found that increased liability for cannabis use disorder is genetically correlated with low educational attainment, early age at first birth, and high body-mass index, traits that show opposite directions of association with lifetime cannabis ever-use. We also found that genetic liability for cannabis use disorder is associated with increased risk of mental health problems, infectious diseases, and respiratory illnesses in a large independent sample. Finally, we found a significant association between increased polygenic liability for cannabis use disorder and low white matter volume in cannabis-naive children, suggesting a potential role of cannabis-related genetic predisposition in early brain development.

Implications of all the available evidence

Cannabis use disorder is a psychiatric illness that is genetically associated with many negative outcomes (including increased risk of psychiatric disorders and respiratory illnesses). Lifetime cannabis use and cannabis use disorder show at least partially divergent genetic influences and associations with relevant traits. Given increasingly permissive cannabis laws and positive perceptions of its safety, the recognition that cannabis use disorder is a serious psychiatric illness should spur prevention and treatment efforts.

Prof O Mors PhD, Prof P B Mortensen, Prof M Nordentoft MD, Prof T Werge PhD, Prof A D Børglum); **Center for Genomics and Personalized Medicine, Aarhus, Denmark** (D Demontis, Prof A D Børglum); **CNS Department** (T E Thorgeirsson PhD, G W Reginsson PhD), **and Statistics Department** (D F Gudbjartsson PhD), **deCODE Genetics/Amgen** (Prof K Stefansson MD), **Reykjavik, Iceland; Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA** (R K Walters PhD, R Wedow PhD); **Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA** (R K Walters, R Wedow); **Division of Human Genetics, Department of Psychiatry,** (R Polimanti PhD, F R Wendt PhD, H Zhou PhD, Prof J Gelernter MD), **and Department of Genetics, and Department of Neuroscience** (Prof J Gelernter), **Yale School of Medicine, New Haven, CT, USA; Veterans Affairs Connecticut Healthcare System, West Haven, CT, USA** (R Polimanti, F R Wendt, H Zhou, Prof J Gelernter); **Department of Psychiatry** (S Sanchez-Roige PhD, Prof S A Brown PhD, Prof T L Wall PhD), **and Department of Psychology and Office of Research Affairs** (Prof S A Brown), **University of California San Diego, La Jolla, CA, USA; Department of Medicine, Division of Genetic Medicine** (S Sanchez-Roige, L K Davis PhD), **and Department of Psychiatry and Behavioral Sciences, and Department of Biomedical Informatics** (L K Davis), **Vanderbilt University Medical Center, Nashville, TN, USA; Department of Psychological and Brain Sciences, Washington University in Saint Louis, St. Louis, MO, USA** (S E Paul BA, R Bogdan PhD); **Division of Psychiatry, University of Edinburgh, Edinburgh, UK** (T-K Clarke PhD, M J Adams PhD, Prof A M McIntosh MD); **Department of Medical and Molecular Genetics, Indiana**

	European ancestry		African ancestry		European ancestry—case-control individuals	
	Cases	Controls	Cases	Controls	Cases	Controls
Case-control studies						
CATS	958	453	..	..	958	453
CADD	397	699	59	55	397	699
CHDS	201	420	..	..	201	420
FSCD	226	314	199	401	226	314
COGEND Nico	306	607	154	313	306	607
COGEND SAGE	228	830	79	187	228	830
GEDI-GSMS	81	491	..	..	81	491
ADAA	..	..	1000	811	..	..
Total	2397	3814	1491	1767	2397	3814
Family-based studies						
BLTS	170	1216	..	..	147	662
MCTFR	449	1625	..	..	389	1185
Yale Penn 1	916	833	1189	1857	839	657
Yale Penn 2	557	497	355	548	557	495
bigCOGA	2206	5053	813	1725	248	795
CEDAR	64	148	..	..	64	148
OZ-ALC	593	4893	..	..	470	1534
VTSABD	99	734	..	..	94	361
IASPSAD	104	613	..	..	84	353
Total	5158	15 612	2357	4130	2892	6190
Summary statistics						
Add health	722	4071	..	..	..	..
PGC studies total	8277	23497	3848	5897	5289	10 004
iPSYCH	2758	53326	0	0	2758	53 326
deCODE	6033	280 396	0	0	6033	280 396
Total (European ancestry)	17 068	357 219	..	..	14 080	343 726
Total (Transancestral)	20 916 cases; 36 3116 controls	..	..	..	..	..

ADAA=Alcohol Dependence in African Americans. BLTS= Brisbane Longitudinal Twin Study. CADD=Center on Antisocial Drug Dependence. CATS=Comorbidity and Trauma Study. CEDAR= Center for Education and Drug Abuse Research. CHDS=Christchurch Health and Development Study. COGA=Collaborative Study on the Genetics of Alcoholism. COGEND=Collaborative Genetic Study of Nicotine Dependence. FSCD=Family Study of Cocaine Dependence. GEDI=Gene-Environment-Development Initiative. GSMS=Great Smoky Mountains Study. IASPSAD=Irish Affected Sib-Pair Study of Alcohol Dependence. MCTFR=Minnesota Center for Twin and Family Research. OZ-ALC=Australian Alcohol and Nicotine Studies. SAGE=Study of Addiction: Genetics and Environment. VTSABD=Virginia Twin Studies of Adolescent Behavioral Development.

Table 1: **Numbers of cases and controls in meta-analysis**

University School of Medicine, Indianapolis, IN, USA (D Lai PhD, L Wetherill PhD, Prof T M Foroud PhD, Prof H J Edenberg PhD); **Department of Psychiatry, University of Pittsburgh Medical Center, Pittsburgh, PA, USA** (D A A Baranger PhD); **School of Engineering and Natural Sciences, Iceland University, Reykjavik, Iceland** (D F Gudbjartsson); **Department of Psychiatry, University of Utah, Salt Lake City, UT, USA** (D E Adkins PhD); **Department**

Methods

Samples

We performed a GWAS of 20 samples in total: 18 from the Psychiatric Genomics Consortium Substance Use Disorders working group (European ancestry 8277 cases, 23 497 controls; African ancestry 3848 cases, 5897 controls), one iPSYCH[9] sample (European ancestry 2758 cases, 53 326 controls), and one deCODE sample (European ancestry 6033 cases, 280 396 controls; table 1; appendix pp 2–8).

This study was approved by the institutional review board at Washington University School of Medicine and was done in accordance with all relevant ethical regulations. Investigators for each contributing study obtained informed consent from their participants and received ethics approvals from their respective review boards in accordance with applicable regulations. Personal identifiers associated with phenotypic information and samples from deCODE were encrypted using a third-party encryption system.[10] The iPSYCH group used pseudonymised unique identifications.[9]

Measures

Psychiatric Genomics Consortium cases met criteria for a lifetime diagnosis of DSM-IV (or DSM-III-R) cannabis abuse or dependence[11] derived from clinician ratings or semi-structured interviews.[7] Cases from the iPSYCH sample had ICD-10 codes of F12.1 (cannabis abuse) or F12.2 (cannabis dependence), or both in the Danish Psychiatric Central Research Register;[12] the remaining individuals in the sample were used as controls. Cases in the deCODE sample met criteria for lifetime DSM-III-R or DSM-IV cannabis abuse or dependence or DSM-5 cannabis use disorder according to diagnoses made at the National Center of Addiction Medicine in Iceland, whereas controls were derived from the general population of Iceland (appendix pp 2–3). Exposure data were not available for some large groups (eg, iPSYCH and deCODE); therefore, controls were defined regardless of lifetime cannabis exposure across all datasets.

Genotyping: quality control and imputation

For the Psychiatric Genomics Consortium, standard procedures for GWAS quality control and imputation were applied using the Ricopili[13] pipeline for case-control groups and the Picopili pipeline for family-based samples. Briefly, variants in each group were filtered for call rate (<5% missingness), followed by individual-level filtering for call rate (<2% missingness) and heterozygosity ($|F^{het}|$ <0·20). If available, chromosome X variants were checked to ensure concordance between genotype sex and reported sex. Variants were then filtered more stringently: variants with more than 2% missingness, differential missingness between cases and controls greater than 2%, invariant markers, and those departing from Hardy-Weinberg equilibrium in cases ($p<1{\cdot}00\times10^{-10}$) or controls ($p<1{\cdot}00\times10^{-6}$) were removed (appendix pp 8–10). Principal components analysis was done on a stringently quality-control set of variants using EIGENSOFT[14,15] to exclude population outliers, infer ancestry among the retained individuals (using the 1000 Genomes Phase 3[16] cosmopolitan reference panel), and derive ancestry-specific principal components for inclusion in analyses (appendix p 10). Sample and variant quality-controlled procedures, including filters for call rate, heterozygosity, and departure from Hardy-Weinberg equilibrium, were done within each ancestry group in each sample. Each group was phased using SHAPEIT[17] and imputed using IMPUTE2[18] to the 1000 Genomes Phase 3[16] cosmopolitan reference panel (appendix pp 10–11). Duplicate

individuals were removed and individuals who were cryptically related across groups were excluded from all but one group (appendix p 12). Single-nucleotide polymorphisms (SNPs) were filtered for INFO score of more than 0·8 and minor allele frequency of at least 0·01 before meta-analysis (appendix pp 13–14).

Quality control of iPSYCH data mirrored the process implemented in the Psychiatric Genomics Consortium, with minor deviations in thresholds for exclusion (appendix p 9).[3] As for deCODE, samples were assayed with several Illumina arrays at deCODE genetics. SNPs with low call rate (<95%), significant deviation from Hardy-Weinberg equilibrium (p<0·001), and excessive inheritance error rates (>0·001) were excluded. We did variant imputation on the basis of the IMPUTE HMM model and long-range phasing.[19] Variants were further filtered for imputation INFO score more than 0·8 and minor allele frequency at least 1% before inclusion in meta-analysis.

Statistical analysis

We did separate association analyses for each sample (ie, 18 individual samples from Psychiatric Genomics Consortium, iPSYCH, and deCODE) by ancestry. For the eight case-control studies from the Psychiatric Genomics Consortium, imputed dosages were analysed using logistic regression models, implemented in the Ricopili pipeline.[13] For family-based samples of the Psychiatric Genomics Consortium, we did association analyses with imputed best-guess genotypes using generalised estimating equations for samples that included only first-degree relatives (eg, sibships), and logistic mixed models for complex pedigrees, in the Picopili pipeline.[7] For calculation of SNP heritability and genetic correlations, subsets of genetically unrelated individuals were selected from each family-based sample from the Psychiatric Genomics Consortium and analysed using logistic regression through Picopili (5289 cases, 10 004 controls). These results were then meta-analysed along with the case-control groups. Psychiatric Genomics Consortium covariates included sex and five to ten within-ancestry principal components to account for population stratification (appendix pp 12–13). Because age was not available in all samples, it was not included as a covariate in the Psychiatric Genomics Consortium analyses. Sensitivity analyses in one representative sample showed this to have no impact on study-specific findings.

In the iPSYCH cohort, logistic regression was done with imputed dosages, covarying for five ancestral principal components, data processing waves, and the presence of another psychiatric disorder (because iPSYCH was established to study major psychiatric disorders, cases of cannabis use disorder and controls include comorbidity).[3] Adding sex as a covariate to iPSYCH analyses has been shown not to alter findings.[20]

deCODE data were analysed using logistic regression of imputed dosage data with sex, age, and county of origin as covariates.[21] To account for inflation due to population stratification and relatedness, test statistics were divided by an inflation factor estimated from linkage disequilibrium score regression (LDSR; appendix p 13).[22]

Effective sample size-weighted meta-analyses across case-control and family-based samples within ancestry were done using METAL (appendix pp 13–14).[23] First, summary statistics of case-control and family-based samples from the Psychiatric Genomics Consortium were combined and weighted by the effective sample size, because effect sizes from case-control logistic regression analyses and family-based analyses using generalised estimating equations and logistic mixed models are not directly comparable. Then the Psychiatric Genomics Consortium results were meta-analysed with those from the iPSYCH and deCODE samples (between-sample genetic correlations [r_g] 0·66–0·70). The summary statistics were filtered such that an SNP had to be present in at least two of the three contributing GWASs (deCODE, iPSYCH, and the Psychiatric Genomics Consortium).

We also did a meta-analysis that excluded related individuals from the family-based samples of the Psychiatric Genomics Consortium, using an inverse variance-weighted scheme, to generate summary statistics that produced effect sizes for use in follow-up analyses (14 080 cases, 343 726 controls). A transancestral meta-analysis using METAL[23] combined results across the European and African ancestry cohorts, comprising 20 916 individuals with cannabis use disorder (17 068 from European ancestry, 3848 from African ancestry) and 363 116 controls (357 219 from European ancestry, 5897 African ancestry; appendix pp 13–14, 17). Conditional analyses were done in GCTA-COJO[24] by conditioning the meta-analysis summary statistics on the lead variants of genome-wide significance.

The FUMA web-based platform[25] version 1.3.5e was used for visualisation and annotation, and MAGMA[26] was used within the FUMA framework to do gene-based association analyses, with SNPs assigned to genes on the basis of physical position (appendix pp 14–15). We also used Hi-C coupled MAGMA to assign non-coding SNPs (intergenic and intronic) to genes on the basis of their chromatin interactions (exonic and promoter SNPs are still assigned to genes on the basis of their genomic location; appendix p 15).[27] Pathway analyses were done using PASCAL to test canonical pathways in the European ancestry sample.[28] All variants within all genes were tested, using default settings, with the structure of linkage disequilibrium estimated using the 1000 Genomes European sample as a reference. We used S-PrediXcan[29] to examine gene expression differences associated with case-control status, using our summary statistics of cannabis use disorder and transcriptome data from the PredictDB Data Repository for 11 brain regions, liver tissue, whole blood, and two types of adipose tissue. We included these tissues because cannabis use disorder is a psychiatric disorder and

of Psychology (A E Adkins PhD, Prof D M Dick PhD), **and Department of Human & Molecular Genetics** (Prof D M Dick, J L Silberg PhD, H H Maes PhD), **College Behavioral and Emotional Health Institute** (A E Adkins), **Virginia Commonwealth University Alcohol Research Center** (J Alexander PhD, S-A Bacanu PhD, B T Webb PhD, Prof B P Riley PhD), **and Department of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics** (R E Peterson PhD, J L Silberg, N A Gillespie PhD, Prof K S Kendler MD, H H Maes), **Virginia Commonwealth University, Richmond, VA, USA; Department of Psychiatry and Behavioral Sciences, SUNY Downstate Health Sciences University, Brooklyn, NY, USA** (T B Bigdeli PhD, J L Meyers PhD, Prof B Porjesz PhD); **Department of Psychological Medicine** (Prof J Boden PhD, Prof J Horwood MSc), **Biostatistics and Computational Biology Unit** (J F Pearson PhD), **and Department of Pathology and Biomedical Science** (J F Pearson, Prof M A Kennedy PhD), **University of Otago, Christchurch, New Zealand; Department for Congenital Disorders, Center for Neonatal Screening, Statens Serum Institut, Copenhagen, Denmark** (J Bybjerg-Grauholm, D M Hougaard); **Department of Molecular, Cellular, and Developmental Biology** (Prof K Krauter PhD), **Department of Integrative Physiology** (M B McQueen ScD), **and Institute for Behavioral Genetics,** (R P Corley PhD, Prof J K Hewitt PhD, Prof M C Stallings PhD), **University of Colorado Boulder, Boulder, CO, USA** (Prof K Krauter); **National Drug and Alcohol Research Centre, University of New South Wales, Sydney, NSW, Australia** (Prof L Degenhardt PhD); **Stanford University Graduate School of Education** (B W Domingue PhD), **and Department of Psychiatry and Behavioral Sciences** (L M Hack MD), **Stanford University, Stanford, CA, USA; Department of Neuroscience, Icahn School of Medicine at**

Mount Sinai, New York, NY, USA (Prof A M Goate DPhil); **QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia** (S D Gordon PhD, P A Lind PhD, Prof N G Martin PhD, Prof S E Medland PhD); **GenOmics, Bioinformatics, and Translational Research Center, Biostatistics and Epidemiology Division, RTI International, Research Triangle Park, Durham, NC, USA** (D B Hancock PhD, Prof E O Johnson PhD); **Brain and Mind Centre, University of Sydney, Sydney, Australia** (Prof I B Hickie MD); **Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN, USA** (J N McClintick PhD, Prof H J Edenberg); **Henri Begleiter Neurodynamics Laboratory, SUNY Downstate Health Sciences University, Brooklyn, NY, USA** (J L Meyers, Prof B Porjesz); **Institute for Molecular Bioscience** (Prof G W Montgomery PhD), **and Queensland Brain Institute** (M J Wright PhD), **University of Queensland, QLD, Australia; Psychosis Research Unit, Aarhus University Hospital, Aarhus, Denmark** (Prof O Mors); **Mental Health Services in the Capital Region of Denmark, Mental Health Center Copenhagen, University of Copenhagen, Copenhagen, Denmark** (Prof M Nordentoft); **School of Pharmacy, University of Pittsburgh, Pittsburgh, PA, USA** (M D Reynolds PhD, Prof R E Tarter PhD, Prof M M Vanyukov PhD); **SAA—National Center of Addiction Medicine, Vogur Hospital, Reykjavik, Iceland** (V Runarsdottir MD, T Tyrfingsson MD); **Department of Medicine (Biomedical Genetics), Boston University School of Medicine, Boston, MA, USA** (R Sherva PhD, Prof L A Farrer PhD); **Institute of Biological Psychiatry, Mental Health Services, Copenhagen University Hospital, Copenhagen, Denmark** (Prof T Werge); **Department of Clinical Medicine, and Center for GeoGenetics, GLOBE Institute, University of Copenhagen, Copenhagen, Denmark** (Prof T Werge); **Department of Psychology, University of Minnesota,**

tetrahydrocannabinol, a key psychoactive cannabis component, accumulates in adipose tissue.[30] Analyses were restricted to the European ancestry meta-analysis because the prediction models were trained on reference transcriptome data from GTEx version 8[31] using only individuals of European ancestry. The significance threshold was corrected for the total number of gene-tissue pairs tested (75 684 tested, $\alpha=6\cdot69\times10^{-7}$).

Heritability explained by common variants (h^2_{SNP}) and genetic correlations with 23 other traits chosen because of previous findings or hypothesised relationships (including cannabis use; appendix pp 15–17) were estimated using LDSR[22,32] on the results of the meta-analysis of case-control individuals of European ancestry (the number of unrelated cases of African ancestry was less than the acceptable sample size threshold for LDSR). Conversion of h^2_{SNP} estimates from observed scale to liability scale was done using a range of estimated population prevalences from 1%[3] to 8·5% (because in some samples we used DSM-IV cannabis abuse or dependence).[33] Significance of genetic correlations with other traits was determined using a Bonferroni correction for 23 tests (including with cannabis use; $\alpha=0\cdot002$). Finally, we examined whether the genetic correlations for cannabis use disorder were significantly different than those for cannabis use[5] using the jackknife procedure implemented through LDSR.[32]

To investigate potential causal relationships, we did latent causal variable analyses on cannabis use disorder and the top genetically correlated traits: educational attainment, age at first birth, Townsend Deprivation Index, smoking initiation, and ADHD (appendix p 16).[34]

We used mtCOJO[35] to condition the summary statistics of cannabis use disorder on loci associated with cannabis use at $p<0\cdot001$[5] to adjust for as many SNPs as possible while retaining computational efficiency. Adjusted summary statistics were used to recompute genetic correlations. Because of the high co-occurrence of cannabis use and tobacco smoking, we also did mtCOJO analyses to condition the summary statistics of cannabis use disorder for loci significantly associated with smoking initiation and cigarettes smoked per day[36] ($p<5\cdot00\times10^{-8}$; excluding 23andMe data, because of restricted access). Moreover, given long-standing interest in the comorbidity of schizophrenia and cannabis misuse, we used mtCOJO to condition the summary statistics of cannabis use disorder on significant schizophrenia loci.[37]

LDSR was used to estimate the genetic correlation between cannabis use disorder and a broad measure of maximum cannabis use frequency. Linear regression was then used to examine the extent to which polygenic risk scores (PRS) for cannabis use disorder predicted a pseudocontinuous measure of self-reported cannabis use frequency, while covarying for age, sex, and 20 ancestral principal components (appendix p 16). PRSice-2[38] was also used to do gene-set enrichment using gene sets and pathways from the Molecular Signatures Database.[39]

PRS for cannabis use disorder were computed using PRS-CS[40] for each of the 66 915 genotyped individuals of European descent in BioVU (appendix pp 16–17). Genotyping and quality control of this sample have been described elsewhere.[41] A logistic regression model was fitted to each of 1335 case or control phenotypes that had at least 100 cases to estimate the odds of each diagnosis given the PRS for cannabis use disorder, after adjustment for sex, median age of the longitudinal electronic health record measurements, and the top ten ancestral principal controls. To explore whether pleiotropic effects of the PRS for cannabis use disorder were mediated by smoking behaviours, we did two phenotype-wide association study (PheWAS) sensitivity analyses: a PheWAS on summary statistics of cannabis use disorder that had been conditioned on the top smoking initiation loci using mtCOJO,[35] and a PheWAS using a diagnosis of tobacco use disorders as an additional covariate in the regression model, which is a conservative over-correction given the extremely high comorbidity expected between cannabis use disorder and tobacco use disorder. We used a Bonferroni-corrected phenome-wide significance threshold of $3\cdot74\times10^{-5}$; this is overly conservative because it incorrectly assumes independence between phenotypes. PheWAS analyses were run using the PheWAS R package, version 0.12.[42]

Data from the Adolescent Brain Cognitive Development Study (Registered; ABCD study)[43] (data release 2.0.1) were used to test the association of PRS for cannabis use disorder with brain structure among 4539 cannabis-naive children (through self-reporting or hair toxicology) of European ancestry (mean age 9·93 years [SD 0·63], 2125 [47%] were girls; appendix p 17). Total bilateral white matter volume, grey matter volume, and intracranial volume were estimated using FreeSurfer 5.3.[44] PRS from the cannabis use disorder GWAS were generated at nine p value thresholds (ie, $p<0\cdot0001$, $p<0\cdot001$, $p<0\cdot01$, $p<0\cdot10$, $p<0\cdot20$, $p<0\cdot30$, $p<0\cdot40$, $p<0\cdot50$, and $p<1\cdot00$), as were PRS for cannabis use.[5] Linear mixed models were used to include scanner (for imaging analyses) and family as nested random effects, done using the lme4 package in R, version 3.6.0. All analyses included as fixed effect covariates the first 20 ancestral principal components, age, sex, age by sex, parents combined income, caregiver education, genotyping batch, caregiver's marital status, prenatal cannabis exposure before and after knowledge of pregnancy, and twin status. Multiple testing within each brain structure phenotype was accounted for by applying random field theory correction across p value thresholds, as this method directly models the overlap across the different PRS thresholds and corrects for the statistical dependence among them.[45]

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full

	Position	SNP	Effect allele	deCODE OR (SE)	deCODE p value	iPSYCH OR (SE)	iPSYCH p value	PGC EUR unrel OR (SE)	PGC EUR unrel p value	PGC EUR comp Z score	PGC EUR comp p value	EUR meta-analysis OR (SE)*	EUR meta-analysis p value*	Trans-ancestral Z score†	Trans-ancestral p value†
Chromosome 7	114 116 881	rs7783012	A	1·10 (0·03)	$5·32\times10^{-4}$	1·09 (0·03)	$2·58\times10^{-3}$	1·11 (0·03)	$9·56\times10^{-5}$	3·47	$5·22\times10^{-4}$	1·11 (0·02)	$1·84\times10^{-9}$	5·97	$2·43\times10^{-9}$
Chromosome 8	27 432 062	rs4732724‡	C	0·90 (0·03)	$3·03\times10^{-4}$	0·84 (0·03)	$5·73\times10^{-8}$	0·98 (0·04)	0·616	–1·91	0·056	0·89 (0·02)	$6·46\times10^{-9}$	–5·95	$2·64\times10^{-9}$

comp=complete meta-analysis (including related individuals and summary statistic cohorts). EUR=European ancestry. OR=odds ratio. PGC=Psychiatric Genomics Consortium. SNP=single nucleotide polymorphism. unrel=unrelated genotyped meta-analysis. *Complete deCODE, iPSYCH, and PGC EUR meta-analysis (excluding related individuals and summary statistic cohorts in the PGC). †Transancestral meta-analysis with deCODE, iPSYCH, and PGC samples (including related individuals and summary statistic cohorts). ‡SNP was only present in half of the PGC samples.

Table 2: **Association statistics for the lead genome-wide significant SNPs across each of the three primary samples (deCODE, iPSYCH, PGC) in the European ancestry and transancestral meta-analyses**

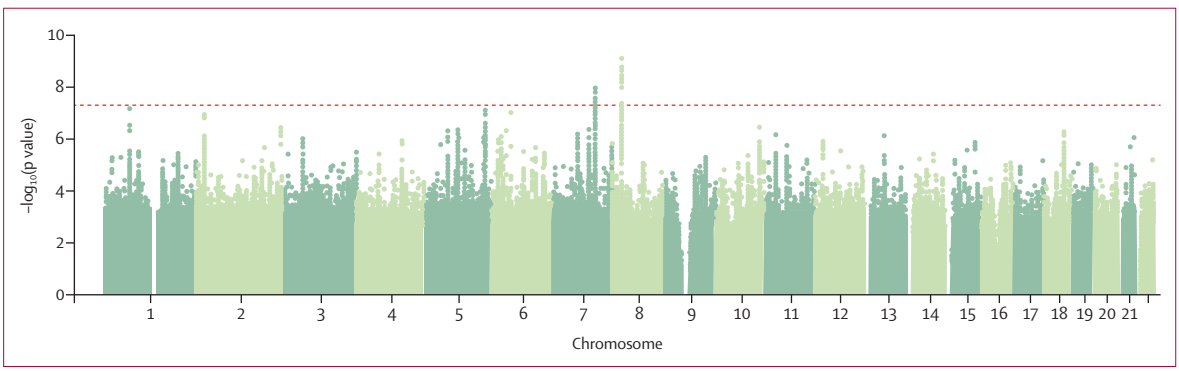

Figure 1: **Manhattan plot of the European ancestry-only genome-wide meta-analysis**

access to all of the data and the final responsibility to submit for publication.

Results

We identified two genome-wide significant loci in the transancestral meta-analysis of cannabis use disorder (African and European ancestries, 20 916 cases, 363 116 controls; appendix pp 17, 20). These loci were significant in the European ancestry meta-analysis but did not reach significance in the much smaller African ancestry analysis (17 068 cases, 357 219 controls *vs* 3848 cases, 5897 controls; table 2). The lead SNPs were rs4732724 on chromosome 8 ($p_{transancestral}=2·64\times10^{-9}$, $p_{European}=6·46\times10^{-9}$, $p_{African}=0·70$) and rs7783012 on chromosome 7 ($p_{transancestral}=2·43\times10^{-9}$, $p_{European}=1·84\times10^{-9}$, $p_{African}=0·09$), with the same direction of effect observed for both ancestries. No additional ancestry-specific loci were observed.

Based on effect sizes and linkage disequilibrium from the case-control European ancestry meta-analysis (cases 14 080, controls 343 726), the genome-wide significant locus on chromosome 8 contains a single association (independent at $R^2<0·1$) with lead SNP rs4732724 (odds ratio [OR] 0·89, 95% CI 0·86–0·93, SE 0·02; $p=6·46\times10^{-9}$; figure 1, appendix pp 21–23). This locus was previously associated with cannabis use disorder in the iPSYCH sample[3] and includes eQTLs for *CHRNA2* (cholinergic receptor nicotinic α2 subunit) in the cerebellum and cerebellar hemisphere and *EPHX2* (epoxide hydrolase 2) in the cerebellum and adipose tissue. One genome-wide significant variant in the chromosome 8 locus (rs1565735) had a CADD score of 13·28, indicating high probability of deleteriousness (appendix p 17). There were additional eQTL signals at this chromosome 8 locus, for *CCDC25* (coiled-coil domain containing 25, in nucleus accumbens, multiple SNPs), *CLU* (clusterin, in adipose, rs2640724), and *STMN4* (stathmin 4, in prefrontal cortex, rs78875955 and rs72477506; appendix p 25).

The chromosome 7 locus is located in an intron of *FOXP2* (Forkhead box protein P2, index SNP, rs7783012; OR 1·11, 95% CI 1·07–1·15, SE 0·02; $p=1·84\times10^{-9}$; figure 1, appendix pp 21–22, 24). The index variant was an eQTL for *FOXP2* in brain (prefrontal cortex, anterior cingulate cortex) and adipose tissue, and demonstrated chromatin interactions with *FOXP2, MDFIC,* and *MIR3666* (appendix p 26).

Inflation in the test statistics (λ=1·10) probably reflects the polygenic architecture of cannabis use disorder, a conclusion supported by LDSR (LDSR intercept 0·99). Conditioning the summary statistics of cannabis use disorder on the lead SNP in each genome-wide significant locus, rs7783012 and rs4732724, did not reveal additional independent significant findings.

Minneapolis, MN, USA (S Zellers MA, Prof W G Iacono PhD, Prof M McGue PhD, S Vrieze PhD); **Institute of Behavioral Science and Department of Sociology, University of Colorado Boulder, Boulder, CO, USA** (Prof J D Boardman PhD); **Department of Psychiatry, University of Vermont Medical Center, Burlington, VT, USA** (Prof W E Copeland PhD); **Department of Sociology, and The Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA** (Prof K M Harris PhD); **Department of Psychiatry, University of Connecticut School of Medicine, Farmington, CT, USA** (Prof V Hesselbrock PhD); **Department of Psychiatry, University of Colorado Denver, Aurora, CO, USA** (Prof C J Hopfer MD); **Center for Studies of Addiction, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA** (Prof H R Kranzler MD); **VISN 4 MIRECC, Crescenz VAMC, Philadelphia, PA, USA** (Prof H R Kranzler); **Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA** (Prof B S Maher PhD); **and Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland** (Prof K Stefansson)

Correspondence to: Dr Emma C Johnson, Department of Psychiatry, Washington University School of Medicine, Saint Louis, MO 63110, USA **emma.c.johnson@wustl.edu**

See **Online** for appendix

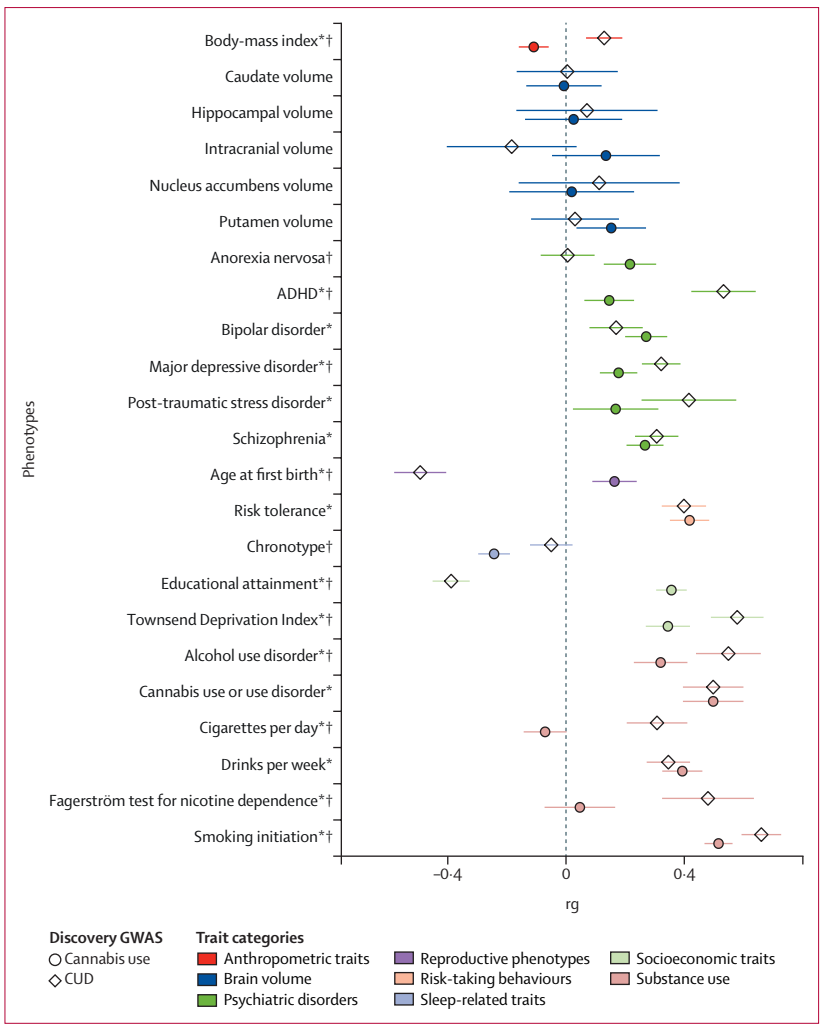

Figure 2: **Genetic correlations between CUD, cannabis use, and other traits of interest**
CUD=cannabis use disorder. GWAS=genome-wide association studies. rg=genetic correlation. *Significantly genetically correlated with CUD. †Significantly different correlations between CUD and cannabis use (α=0·002).

For the **Picopili pipeline** see https://github.com/Nealelab/picopili

For the **PredictDB Data Repository** see http://predictdb.org

For **GWAS of cannabis use frequency** see see https://github.com/Nealelab/UK_Biobank_GWAS

For the **Adolescent Brain Cognitive Development Study** see https://abcdstudy.org/

The gene-wise association analysis of European ancestry summary statistics identified three significant genes (α=2·664×10^{-6}): *FOXP2* (p=7·31×10^{-8}), *PDE4B* (p=6·66×10^{-7}), and *ENO4* (p=3·51×10^{-8}; appendix p 17, 27). No pathways were significant (appendix p 18). Three genes, *NAT6* (amygdala, cortex, frontal cortex), *HYAL3* (both adipose tissues, whole blood, cerebellum, frontal cortex, hippocampus, nucleus accumbens, and spinal cord), and *IFRD2* (cerebellum) were significantly related to cannabis use disorder through genetically regulated gene expression (appendix pp 18, 28). Connecting SNPs to genes via chromatin interaction data revealed significant associations in adult brain tissue (ten genes), fetal brain tissues (12 genes), iPSC-derived astrocytes (11 genes), and iPSC-derived neurons (eight genes); these genes included *HYAL3, ENO4, CHRNA2,* and *FOXP2* (appendix pp 18, 29).

The SNP-heritability (h^2_{SNP}) for cannabis use disorder was 0·067–0·121 (SE 0·006–0·011) on the liability scale, depending on the estimated population prevalence and h^2_{SNP} 0·02 (SE 0·002) on the raw scale. Cannabis use disorder showed significant r_g with 16 of the 23 studied phenotypes, for which the strongest relationships were observed with smoking initiation[36] (r_g 0·66, p=3·20×10^{-83}), Townsend Deprivation Index (a measure of regional poverty[46]; r_g 0·58, p=3·30×10^{-37}), educational attainment[47] (r_g –0·39, p=6·70×10^{-34}), and age at which first child is born (r_g –0·49; p=5·40×10^{-28}; figure 2, appendix p 18). Thus, increased risk of cannabis use disorder is genetically correlated with increased liability for smoking initiation, living in an area of high material poverty, having children at an early age, and low levels of educational attainment. Liability to cannabis use disorder was also positively genetically correlated with alcohol use,[36] nicotine dependence,[48] psychiatric disorders (eg, ADHD,[20] schizophrenia,[37] major depression),[49] and body-mass index (BMI).[50]

The r_g between cannabis use and cannabis use disorder was 0·50 (SE 0·05, p=1·50×10^{-21}). Of the eight genome-wide significant SNPs associated with cannabis use, only four had p<0·05 in the meta-analysis of cannabis use disorder (modest sample overlap between the two studies: genetic covariance intercept 0·014 [SE 0·005]).[5] Conditioning the summary statistics of cannabis use disorder for loci associated with cannabis use neither substantially modified the effect sizes of the genome-wide significant loci (rs4732724, β=–0·11, SE 0·02, p=8·25×10^{-9}; rs7783012, β=0·10, SE 0·02, p=2·62×10^{-9}) nor identified additional novel loci (appendix p 18). The heritability of cannabis use disorder adjusted for cannabis use loci (using mtCOJO[35]) was 0·095 (SE 0·01) on the liability scale (estimated population prevalence 8·5%).

The r_gs with cannabis use disorder and cannabis use were significantly different for 12 of the 22 traits compared (figure 2, appendix p 18). Cannabis use[5] and cannabis use disorder were positively genetically correlated with liability to smoking initiation, schizophrenia, major depressive disorder, risk tolerance, and the Townsend Deprivation Index. Cannabis use[5] was positively genetically correlated with educational achievement and later age at birth of first child, and negatively with BMI. In contrast, cannabis use disorder was genetically correlated with low education attainment, early age at birth of first child, and high BMI. Liability to cannabis use disorder was genetically correlated with nicotine dependence (r_g 0·48, p=1·35×10^{-9}), whereas the genetic correlation of this trait with cannabis use was not significant (p=0·44). In contrast, cannabis use was significantly genetically correlated with chronotype (r_g –0·24, p=6·40×10^{-19}), whereas cannabis use disorder showed no significant correlation with this trait (p=0·18). Conditioning the r_g of cannabis use disorder on cannabis use loci (with p<0·001) made little difference in the magnitude of the r_gs (appendix p 18).

We found no evidence of genetically causal relationships between liability to cannabis use disorder and to any of the most highly correlated traits (ie, educational attainment, age at first birth, Townsend Deprivation

***Figure 3*: PheWAS associations between polygenic risk for CUD and phenotypes in the BioVU biobank**
The 46 phenotypes shown are significantly associated with CUD ($p<3{\cdot}74\times10^{-5}$, corrected for 1335 phenotypes tested). CUD=cannabis use disorder. PheWAS=phenotype-wide association study. NOS=not otherwise specified. SIRS=systemic inflammatory response syndrome.

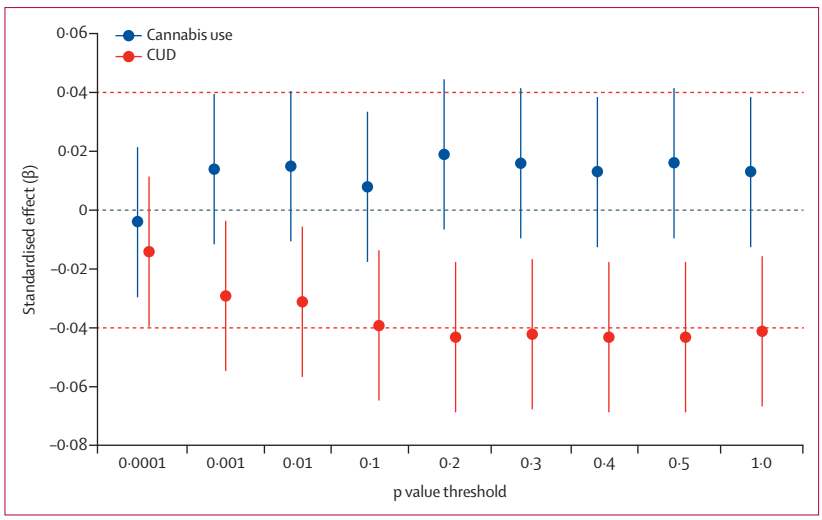

Figure 4: **Polygenic risk score associations with white matter volume in drug-naive children**
Total white matter volume was regressed on polygenic risk scores for CUD and cannabis use (in separate models). CUD=cannabis use disorder.

Index, smoking initiation, or ADHD; genetic causality proportion 0·05–0·27, p=0·128–0·856; appendix p 18).

Liability to cannabis use disorder and maximum cannabis use frequency in the UK Biobank were genetically correlated (r_g 0·75, p=1·80×10^{-6}). PRS for cannabis use disorder were significantly associated with our pseudocontinuous measure of cannabis use frequency in the UK Biobank (maximum R^2 0·04%, Z 7·42, p=1·15×10^{-13}, threshold p<0·3; appendix p 18, 30). 65 of 12 461 gene sets and pathways were significantly enriched, highlighting involvement of CNS morphogenesis (transcription factor Nkx–2·2 target genes, R^2 0·02%, Z 4·46, p=8·22×10^{-6}) and immune responses to exogenous compounds (ZFP91 target genes R^2 0·01%, Z 4·41, p=1·01×10^{-5}; $CD4^+$ T-cell R^2 0·02%, Z 4·41, p=3·79×10^{-6}; and macrophage gene sets R^2 0·01%, Z 4·62, p=1·04×10^{-5}; appendix p 18).

Of 1335 phenotypes in the BioVU biobank, 46 were significantly associated with the PRS for cannabis use disorder (p<3·74×10^{-5}; figure 3, appendix p 18). The phenotype groups with the most abundant associations were mental disorders (n=12), the strongest associations being with tobacco use disorder (cases 5280, OR 1·18, 95% CI 1·13–1·23, SE 0·02; p=2·66×10^{-27}) and substance use disorders (cases 6155, OR 1·18, SE 0·01, 95% CI 1·16–1·20; p=1·24×10^{-30}), mood disorders (cases 9588, OR 1·09, SE 0·01, 95% CI 1·07–1·11; p=2·38×10^{-12}) and suicidal ideation or attempt (cases 689, OR 1·27, SE 0·04, 95% CI 1·17–1·37; p=1·81×10^{-9}); respiratory diseases (n=12), such as respiratory failure (cases 4485, OR 1·11, SE 0·02, 95% CI 1·07–1·15; p=4·45×10^{-10}) or chronic airway obstruction (cases 4436, OR 1·13, SE 0·02, 95% CI 1·09–1·18; p=5·64×10^{-14}); endocrine or metabolic conditions (n=3), such as disorders of fluid (cases 12 562, OR 1·06, SE 0·01, 95% CI 1·04–1·08; p=5·77×10^{-8}); infectious diseases (n=4), such as viral hepatitis (cases 135, OR 1·3, SE 0·03, 95% CI 1·23–1·38; p=3·34×10^{-20}); and digestive diseases (n=3), including cirrhosis of liver (cases 1928, OR 1·14, SE 0·02, 95% CI 1·10–1·19; p=2·49×10^{-8}).

A secondary pheWAS analysis in BioVU using summary statistics of cannabis use disorder conditioned on smoking initiation revealed attenuated findings, with only ten codes now passing Bonferroni corrections; anxiety disorder, viral hepatitis, and several respiratory codes were still significant. When we conditioned the pheWAS on tobacco use disorder, some associations remained significant (respiratory conditions, viral hepatitis), whereas other associations (eg, anxiety disorder) were no longer associated with PRS for cannabis use disorder (appendix p 18).

The PRS for cannabis use disorder were significantly associated with reduced total white matter volume in cannabis-naive children from the ABCD Study (standardised β=–0·04; p=0·001; figure 4), explaining up to 0·18% of the variance in white matter volume at the most predictive threshold of p<0·5 (appendix p 18). Children in the highest quartile of PRS, on average, had a white matter volume that was 1% lower than those in the lowest quartile. Results remained significant when including intracranial volume as a covariate (standardised β=–0·08, p=0·01) and when excluding 1246 (27%) of 4539 children who used any substance (standardised β=–0·05, p=0·001), or when excluding 2482 (54%) of 4539 who used any substance or were prenatally exposed to any substance (standardised β=–0·05, p=0·03). The PRS for cannabis use were not significantly correlated with white matter volume (figure 3). After correction for multiple testing, there was no association between PRS for cannabis use disorder or cannabis use and grey matter volume (all p>0·01; appendix p 18, 31).

Discussion

This GWAS meta-analysis confirmed one previously identified locus on chromosome 8 as associated with cannabis use disorder and identified a new locus on chromosome 7. The lead variant at the chromosome 7 locus (rs7783012) is a *cis*-eQTL for *FOXP2* expression in brain and adipose tissue. *FOXP2* was also significantly implicated in gene-based tests that incorporated information about chromatin interactions in iPSC-derived astrocytes (appendix p 29). rs7783012 has also been associated with measures related to externalising behaviours (eg, ADHD,[20] age at first sexual intercourse,[51] generalised risk tolerance)[52] and with educational attainment.[47] *FOXP2* is essential to synaptic plasticity and has been implicated in the normal development of speech and language acquisition[53] However, because of the prominence of the protein product of *FOXP2* as a regulator of numerous genes, indirect pathways of vulnerability beyond risk-taking are also possible.

Individual SNPs on chromosome 8 are eQTLs for *CHRNA2* and *EPHX2*, extending previous work by

Demontis and colleagues[3] in iPSYCH, with replication in the deCODE data. Note that iPSYCH and deCODE are the main contributors to this finding in the meta-analysis ($p_{iPSYCH}=5{\cdot}73\times10^{-8}$, $p_{deCODE}=0{\cdot}0003$, $p_{PGC}=0{\cdot}06$; appendix p 23). A large GWAS of schizophrenia[54] has also implicated this variant ($p=3{\cdot}68\times10^{-6}$), but conditioning for top schizophrenia loci did not modify the association with cannabis use disorder ($p=4{\cdot}33\times10^{-8}$; appendix p 18). Given the role of *CHRNA2* variants in tobacco smoking,[36] it is plausible that the findings for cannabis use disorder and schizophrenia are partially driven by the high rates of tobacco use in those populations.[55] However, conditioning cannabis use disorder on the GWAS of cigarettes per day increased the significance of the lead variant rs4732724 ($p=4{\cdot}16\times10^{-9}$; appendix p 18), although a different SNP was identified as the lead SNP (rs11783093). When rs11783093 was conditioned for the GWAS of smoking initiation, the signal was attenuated ($p=1{\cdot}55\times10^{-6}$; appendix p 18). These findings suggest that the chromosome 8 signal might be partly driven by smoking initiation, or indicative of a pleiotropic effect with a stronger impact on cannabis use disorder than on smoking initiation.[36] Despite the plausibility of *CHRNA2* in the cause of cannabis use disorder, it is worth noting that *EPHX2*, which is involved in the metabolism of cannabinoids,[56–58] was also identified in eQTL analyses but not supported by other post-hoc analyses (appendix p 29).

Cannabis use and cannabis use disorder were modestly genetically correlated (r_g 0·50) but conditioning for cannabis use loci did not substantially reduce the heritability of cannabis use disorder, and although it reduced the significance of the top loci, the effect sizes remained consistent. Although this does not fully account for possible index-event bias,[59] it suggests that the findings are not due to cannabis exposure alone. Cannabis use and cannabis use disorder also show divergent genetic relationships with educational attainment,[47] BMI,[50] and age at birth of first child, with cannabis use disorder indexing greater impairment in these psychosocial and anthropometric indices than cannabis use. This divergence is similar to that found between alcohol intake and alcohol use disorder.[7,8]

We found genetic overlap between cannabis use disorder and several mental health phenotypes, respiratory illnesses, and infectious diseases in the BioVU biobank. The strongest association was with tobacco use disorder, but conditioning for loci associated with smoking initiation retained many of the pheWAS associations at significant levels, including anxiety, phobic and dissociative disorders, respiratory failure, and viral hepatitis. An even more stringent analysis that covaried for tobacco use disorder revealed independent associations with viral hepatitis, type 1 diabetes, respiratory measures, and pain, but not mental health. These associations could reflect genuine pleiotropy (eg, with risk-taking behaviours and injection drug use) or index putatively causal peripheral effects of cannabis. Cannabis use frequency in the UK Biobank was genetically correlated with cannabis use disorder as well, but, similarly to other psychiatric and behavioural traits,[60] the PRS for cannabis use disorder explained only a small proportion of variance in cannabis use frequency (R^2 0·04%).

Some previous cross-sectional studies have linked differences in grey matter volume with cannabis use and dependence;[61] however, a large mega-analysis did not find reductions in global or regional volumes in cannabis-dependent adults compared with controls.[62] In our study, the association between PRS for cannabis use disorder and white matter volume persisted in the subset of children who were not exposed to any substance, including prenatally. This finding suggests that polygenic liability for cannabis use disorder might index differences in white matter volume in the developing brain, independently of the onset of substance use behaviours. Still, the association between PRS for cannabis use disorder and white matter was small (R^2 0·15–0·18%), and additional studies are needed to confirm this association.

Some limitations are noteworthy. Our African ancestry sample was under-powered; more data are needed, particularly in light of potential disparities that result from a majority of genetic studies focusing on European-ancestry populations.[63,64] We had little or no information regarding comorbid psychiatric disorders for the majority of PGC samples; however, we did conditional analyses to account for these and it made little difference. Information regarding lifetime cannabis exposure and the potency of cannabis used was scarce. Our estimates of genome-wide SNP-h^2 were far lower than the h^2 estimated from twin and family studies (0·07–0·12 *vs* 0·5–0·7). This discrepancy between pedigree-estimated heritability and SNP-heritability is common across essentially all substance use disorders, and might be due to low power, some heritability residing in variants too rare to be included in our GWAS, and insufficient coverage of optimal common-variant genomic coverage in available microarray data even after imputation. An additional limitation is that we did not do formal Mendelian randomisation[65] analysis. To do this analysis, we would have needed to remove sample overlap between our cannabis use disorder GWAS and the other GWASs of interest, which would have greatly decreased our statistical power. However, after doing latent causal variable analyses,[34] an approach related to mendelian randomisation that can account for sample overlap among the input GWAS, there was no significant evidence of causal relationships between liability to cannabis use disorder and to any of the top genetically correlated traits: educational attainment, age at first birth, Townsend Deprivation Index, smoking initiation, or ADHD (appendix p 16). Overall, estimates of genetic overlap might also be sensitive to sample characteristics, such as older volunteers in the UK Biobank cohort[66] and some younger registry-based cohorts in our GWAS. In

addition, imbalance between cases and controls could have affected our findings, although we did not observe substantial genetic heterogeneity (appendix pp 23–24).

In conclusion, our findings provide further evidence that cannabis use disorder is a serious, psychiatric illness with genetic and neurobiological influences that diverge at least partially from cannabis use. From a public health perspective, the recognition that cannabis use disorder is a serious form of psychopathology should spur efforts to identify and aid at-risk individuals in the face of escalating cannabis use worldwide, especially among adolescents.

Contributors

SAB, DD, ADB, JHo, GWM, MN, AA, RB, HJE, IBH, JG, ECJ, TET, and RKW designed the study. MJA, SAB, JDB, AMG, AMM, HRK, DL, BTW, RET, S-AB, VH, BSM, AEA, KSK, WEC, ADB, FRW, SEM, JB-G, LD, NGM, DBH, JHo, RPC, LW, MAK, DMD, JPR, MDR, MCS, MMV, BPR, LJB, LF, JB, KKB, JG, NAG, RAG, DFG, KMH, SMH, ACH, JKH, IBH, DMH, WGI, EOJ, KK, PAFM, HHM, MM, MBM, GWM, OM, PBM, ECN, JFP, BP, VR, NLS, RS, JLS, TET, TT, SV, TLW, TW, MJW, and RW collected and interpreted the data. DAAB, AMM, RP, ASH, IBH, DD, RET, BSM, FRW, SEM, TBB, SEP, LMH, DBH, LW, MAK, DEA, SZ, JA, GWR, AA, RB, SS-R, T-KC, BWD, NAG, SDG, DFG, SMH, JHe, ECJ, EOJ, REP, PAL, GWM, TET, RKW, and RW analysed the data. MJA, DAAB, DL, RP, ASH, DD, S-AB, DBH, DEA, GWR, AA, RB, SS-R, JDB, LKD, HJE, JG, DFG, JHo, JHe, ECJ, NLS, TET, RKW, and HZ gave the statistical input. ASH, ADB, FRW, SEP, JLM, AA, RB, T-KC, HJE, JG, ECJ, TET, and RKW wrote the manuscript. SAB, AMG, HRK, DL, LAF, RP, ASH, DD, RET, WEC, ADB, SEP, DBH, JHo, RPC, LW, MAK, DMD, MCS, MMV, GWR, AA, RB, JNM, SS-R, JDB, KKB, LKD, HJE, JG, ECJ, ECN, NLS, TET, RKW, and HZ edited the manuscript. AMM, KSK, SEM, LD, DMD, JLM, BPR, AA, RB, HJE, JG, NAG, ACH, EOJ, HHM, ECN, and TET provided phenotype expertise. CJH, AMM, LAF, ADB, KSK, NGM, JHo, AA, RB, LKD, HJE, TMF, JG, KMH, JKH, WGI, HHM, ECN, BP, KS, TET, SV, RKW, TW, and MJW supervised the study.

Declaration of interests

TW has acted as an advisor and a lecturer to H Lundbeck A/S. LJB and the spouse of NLS are listed as inventors on Issued US Patent 8 080 371, Markers for Addiction, covering the use of particular SNPs in determining the diagnosis, prognosis, and treatment of addiction. AMM has received research support from Eli Lilly, Janssen, Pfizer, and the Sackler Foundation. HRK is a member of the American Society of Clinical Psychopharmacology's Alcohol Clinical Trials Initiative, which was supported in the last 3 years by AbbVie, Alkermes, Ethypharm, Indivior, Lilly, Lundbeck, Otsuka, Pfizer, Arbor, and Amygdala Neurosciences. HRK and JG are named as inventors on PCT patent application number 15/878 640, entitled Genotype-guided dosing of opioid agonists, filed Jan 24, 2018. LD reports untied educational grant funding to research studies of new opioid medications in Australia from Indivior, Mundipharma, Seqirus, and Reckitt Benckiser.

Data sharing

Summary statistics will be made available for download at https://www.med.unc.edu/pgc/download-results/.

Acknowledgments

The Psychiatric Genomics Consortium's Substance Use Disorders working group is supported by MH109532 with funding from the National Institute of Mental Health (NIMH) and the National Institute on Drug Abuse (NIDA). We gratefully acknowledge previous support from the National Institute on Alcohol Abuse and Alcoholism (NIAAA) and thank all our contributing investigators and study participants who make this research possible. We acknowledge Dr E Jane Costello for leadership in collecting and characterising the Gene-Environment-Development Initiative–Great Smoky Mountains Study sample and her valuable insights on the development of this study and interpretation of findings. This research was done using the UK Biobank Resource, application numbers 4844 and 58146. Individual funding was provided by the following grants: MH109532 (AA, HJE, JG); F32AA027435 (ECJ); K02DA032573 (AA); 1U01MH109514-01 (ADB); 5T32DA007261 (AH); T32MH018951 (DAAB); NIH F32 MH122058 (FRW); Center for Genomics and Personalized Medicine and the Centre for Integrative Sequencing, iSEQ, Aarhus University, Denmark (ADB); the European Commission, Horizon 2020, grant number 667302 (ADB); R01-DA034076 (TET); R01HD060726 (BWD, KMH, JDB, MBM); Health Research Council of New Zealand 16/600 (JB, JHo, MAK, JFP); Health Research Council of New Zealand 11/792 (JB, JHo); AA027827, AG061162 (RB); U24DA041147 (SAB); R01HD093651, R01MH117559, P30DA023026 (WEC); DA011015, DA038504 (RPC); K02 AA018755 (DMD); P01-HD031921, R01-HD073342 (KMH);R01 DA018673, DA025109, DA024413, DA016977 (HHM); Wellcome Trust (104036/Z/14/Z, 216767/Z/19/Z), UKRI MRC (MC_PC_17209, MR/S035818/1) (AMM); K01MH113848, The Brain & Behavior Research Foundation NARSAD grant 28632 (REP); R21 DA047527, R21 DC018098 (RP); SAMHSA Grant # 1H79TI081668 (MDR); R01DA026911, R21DA046791 (NLS); R01HD050735, U54EB020403 (subaward), NHMRC Australia (496682; 1009064) (MJW); Tobacco-Related Disease Research Program of the University of California Grant Number T29KT0526,Families for Borderline Personality Disorder Research (Beth and Rob Elliott) 2018 NARSAD Young Investigator Grant #27676 (SS-R). The iPSYCH team was supported by grants from the Lundbeck Foundation (R165–2013–15320, R102-A9118, R155–2014–1724, and R248–2017–2003), the European Commission (Horizon 2020, grant number 667302), and the universities and university hospitals of Aarhus and Copenhagen, Denmark. The Danish National Biobank resource was supported by the Novo Nordisk Foundation. Data handling and analysis on the GenomeDK HPC facility was supported by NIMH (1U01MH109514–01 to ADB). High-performance computer capacity for handling and statistical analysis of iPSYCH data on the GenomeDK HPC facility was provided by the Center for Genomics and Personalized Medicine and the Centre for Integrative Sequencing, iSEQ, Aarhus University, Denmark (grant to ADB). Work at deCODE genetics was partly funded by the NIDA (R01–DA034076). The BioVU dataset(s) used for the PheWAS analyses described were obtained from Vanderbilt University Medical Center's BioVU, which is supported by numerous sources: institutional funding, private agencies, and federal grants. These grants include the NIH funded Shared Instrumentation Grant S10RR025141, and Clinical Translational Sciences Award grants UL1TR002243, UL1TR000445, and UL1RR024975. Genomic data are also supported by investigator-led projects that include U01HG004798, R01NS032830, RC2GM092618, P50GM115305, U01HG006378, U19HL065962, and R01HD074711, and additional funding sources https://victr.vumc.org/biovu-funding/]. Dr Degenhardt is funded by the NIH (R01 DA144740) and the Australian National Health and Medical Research Council Research (1135991). Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development (ABCD) Study, held in the NIMH Data Archive. The ABCD Study is supported by the NIH and additional federal partners under award numbers U01DA041022, U01DA041028, U01DA041048, U01DA041089, U01DA041106, U01DA041117, U01DA041120, U01DA041134, U01DA041148, U01DA041156, U01DA041174, U24DA041123, U24DA041147, U01DA041093, and U01DA041025. A listing of participating sites and a complete listing of the study investigators can be found online. ABCD consortium investigators designed and implemented the study or provided data, or both, but did not necessarily participate in analysis or writing of this study. This manuscript reflects the views of the authors and might not reflect the opinions or views of the NIH or ABCD consortium investigators. The ABCD data repository grows and changes over time. The ABCD data used in this study came from the NIMH Data Archive 2.0.1. Funding support for the Comorbidity and Trauma Study (dbGAP accession number, phs000277.v1.p1) was provided by NIDA (R01 DA17305); GWAS genotyping services at the Center for Inherited Disease Research at The Johns Hopkins University were supported by NIH (contract N01-HG-65403). Funding support for the Center for Education and Drug Abuse Research (dbGAP accession number, phs001649.v1.p1) was provided by NIDA (P50 DA005605).
The Christchurch Health and Development Study has been supported by funding from the Health Research Council of New Zealand, the National Child Health Research Foundation (Cure Kids), the Canterbury Medical Research Foundation, the New Zealand Lottery Grants Board, the University of Otago, the Carney Centre for

For the **additional funding sources** see https://victr.vumc.org/biovu-funding/

For the **full list of supporters** see https://abcdstudy.org/federal-partners.html

For the **ABCD consortium investigators** see https://abcdstudy.org/scientists/workgroups/

Pharmacogenomics, the James Hume Bequest Fund, US NIH grant MH077874, and NIDA grant, a developmental model of gene-environment interplay in substance use disorders (R01DA024413) 2007–12. The Collaborative Study on the Genetics of Alcoholism (COGA; Principal Investigators B Porjesz, V Hesselbrock, and T Foroud; Scientific Director, A Agrawal; Translational Director, D Dick) includes 11 different centres: University of Connecticut (V Hesselbrock); Indiana University (H J Edenberg, T Foroud, J Nurnberger Jr, Y Liu); University of Iowa (S Kuperman, J Kramer); SUNY Downstate (B Porjesz, J Meyers, C Kamarajan, A Pandey); Washington University in St Louis (L Bierut, J Rice, K Bucholz, A Agrawal); University of California at San Diego (M Schuckit); Rutgers University (J Tischfield, A Brooks, R Hart); The Children's Hospital of Philadelphia, University of Pennsylvania (L Almasy); Virginia Commonwealth University (D Dick, J Salvatore); Icahn School of Medicine at Mount Sinai (A Goate, M Kapoor, P Slesinger); and Howard University (D Scott). Other COGA collaborators include L Bauer (University of Connecticut); L Wetherill, X Xuei, D Lai, S O'Connor, M Plawecki, S Lourens (Indiana University); L Acion (University of Iowa); G Chan (University of Iowa; University of Connecticut); D B Chorlian, J Zhang, S Kinreich, G Pandey (SUNY Downstate); M Chao (Icahn School of Medicine at Mount Sinai); A Anokhin, V McCutcheon, S Saccone (Washington University); F Aliev, P Barr (Virginia Commonwealth University); and H Chin, and A Parsian (NIAAA Staff Collaborators). We continue to be inspired by our memories of Henri Begleiter and Theodore Reich, founding Principal Investigator and Co-Principal Investigator of COGA, and also owe a debt of gratitude to other past organisers of COGA, including Ting-Kai Li, P Michael Conneally, Raymond Crowe, and Wendy Reich for their crucial contributions. This national collaborative study is supported by NIH Grant U10AA008401 from NIAAA and NIDA. We thank Kim Doheny and Elizabeth Pugh from the Center for Inherited Disease Research and Justin Paschall from the NCBI dbGaP staff for valuable assistance with genotyping and quality control in developing the dataset available at dbGaP (phs000125.v1.p1, phs000763.v1.p1, phs000976.v1.p1). Support for the Study of Addiction: Genetics and Environment (SAGE) was provided through the NIH Genes, Environment and Health Initiative (GEI; U01 HG004422; dbGaP study accession phs000092.v1.p1). SAGE is one of the GWAS funded as part of the Gene Environment Association Studies (GENEVA) under GEI. Assistance with phenotype harmonisation and genotype cleaning, as well as with general study coordination, was provided by the GENEVA Coordinating Center (U01 HG004446). Assistance with data cleaning was provided by the National Center for Biotechnology Information. Support for collection of datasets and samples was provided by COGA (U10 AA008401), the Collaborative Genetic Study of Nicotine Dependence (P01 CA089392; see also phs000404.v1.p1), and the Family Study of Cocaine Dependence (R01 DA013423, R01 DA019963). Funding support for genotyping, which was done at the Johns Hopkins University Center for Inherited Disease Research, was provided by the NIH GEI (U01HG004438), NIAAA, NIDA, and the NIH contract High throughput genotyping for studying the genetic contributions to human disease (HHSN268200782096C). The Great Smoky Mountains Study project (phs000852.v1.p1) was supported by NIDA (U01DA024413, R01DA11301), NIMH (R01MH063970, R01MH063671, R01MH048085, K01MH093731, and K23MH080230), NARSAD, and the William T Grant Foundation. We are grateful to all the Great Smoky Mountains Study and Caring for Children in the Community study participants who contributed to this work. The following grants supported data collection and analysis of Center on Antisocial Drug Dependence (dbGAP in progress): DA011015, DA012845, DA021913, DA021905, DA032555, and DA035804. Alcohol Dependence in African Americans was funded by NIH grant R01 AA017444. Brisbane Longitudinal Twin Study was supported by the US NIDA (R00DA023549), and by the Australian Research Council (DP0343921, DP0664638, 464914, 619667, FT110100548). Gene-Environment-Development Initiative Virginia Commonwealth University (VTSABD; dbGAP in progress) was supported by NIDA (U01DA024413, R01DA025109), NIMH (R01MH045268, R01MH055557, R01MH068521), and the Virginia Tobacco Settlement Foundation, grant 8520012. We are grateful to all the Virginia Twin Studies of Adolescent Behavioral Development, Young Adult Follow-Up, Transitions to Substance Use study participants who contributed to this work. Minnesota Center for Twin and Family Research (phs000620.v1.p1) contributing to this publication was supported by the NIH under award numbers DA005147, DA013240, DA024417, DA036216, AA009367, MH066140, DA042755, DA037904, HG008983, and DA044283. Funding for Substance Use Disorder Liability: Candidate System Genes study was supported by R01 DA019157 and P50 DA005605. Yale-Penn (phs000425.v1.p1; phs000952.v1.p1) was supported by NIH grants RC2 DA028909, R01 DA12690, R01 DA12849, R01 DA18432, R01 AA11330, and R01 AA017535 and the Veterans Affairs Connecticut and Philadelphia Veterans Affairs Mental Illness Research, Educational, and Clinical Centers. Australian Alcohol and Nicotine studies (OZ-ALC-NAG; phs000181.v1.p1) were supported by NIH grants AA07535,AA07728, AA13320, AA13321, AA14041, AA11998, AA17688, DA012854, and DA019951, by grants from the Australian National Health and Medical Research Council (241944, 339462, 389927,389875, 389891, 389892, 389938, 442915, 442981, 496739, 552485, and 552498), by grants from the Australian Research Council (A7960034, A79906588, A79801419, DP0770096, DP0212016, and DP0343921); and by the 5th Framework Programme (FP-5) GenomEUtwin Project (QLG2-CT-2002–01254). GWAS genotyping at Center for Inherited Disease Research was supported by a grant to the late Richard Todd, MD, PhD, former Principal Investigator of grant AA13320. Irish Affected Sib-Pair Study of Alcohol Dependence GWAS data collection and analysis was supported by NIAAA grants P20-AA-017828 and P50-AA-022537. Sample collection was supported by R01-AA-011408. Control genotyping was supported by NIMH grant R01-MH-083094 and Wellcome Trust Case Control Consortium 2 grant WTCCC-084710. This research uses data from Add Health, a programme project directed by Kathleen Mullan Harris and designed by J Richard Udry, Peter S Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. No direct support was received from grant P01-HD31921 for this analysis. We additionally thank the groups who directly shared GWAS results. We would like to acknowledge all participating groups of the International Cannabis Consortium, and in particular the members of the working group including Sven Stringer, Camelia Minica, Karin Verweij, Hamdi Mbarek, Eske Derks, Nathan Gillespie, and Jacqueline Vink. We also wish to thank the ENIGMA consortium for providing GWAS results on subcortical brain volumes. Finally, we acknowledge the valuable contribution of groups who have publicly released summary statistics from their respective GWAS. Specifically, thanks to researchers from Schumann and colleagues (2016) including the CHARGE+ and AlcGen consortia and to all members of Psychiatric Genomics Consortium (PGC). Individual-level data from the genotyped cohorts and cohort-level summary statistics will be made available to researchers following an approved analysis proposal through the PGC Substance Use Disorder group with agreement of the cohort Principal Investigators; contact the corresponding authors for details. Some cohort data are also available from dbGaP except when prohibited by Institutional Review Board, funding requirements or European Union data restrictions (accession numbers: phs000277.v1.p1, phs000092.v1.p1, phs000852.v1.p1, phs000404.v1.p1, phs000092.v1.p1, phs000620.v1.p1, phs001649.v1.p1, phs000425.v1.p1, phs000181.v1.p1, phs000763.v1.p1, phs000125.v1.p1).

For the **Add Health data files** see http://www.cpc.unc.edu/addhealth

For the **subcortical brain volume results** see http://enigma.ini.usc.edu/research/download-enigma-gwas-results/

For **CHARGE+ and AlcGen consortia** see https://grasp.nhlbi.nih.gov/FullResults.aspx

For the **Psychiatric Genomics Consortium** see http://www.med.unc.edu/pgc/results-and-downloads

References

1 Verweij KJH, Zietsch BP, Lynskey MT, et al. Genetic and environmental influences on cannabis use initiation and problematic use: a meta-analysis of twin studies. *Addiction* 2010; **105:** 417–30.

2 Agrawal A, Lynskey MT, Hinrichs A, et al. A genome-wide association study of DSM-IV cannabis dependence. *Addict Biol* 2011; **16:** 514–18.

3 Demontis D, Rajagopal VM, Thorgeirsson TE, et al. Genome-wide association study implicates CHRNA2 in cannabis use disorder. *Nat Neurosci* 2019; **22:** 1066–74.

4 Sherva, R, Wang Q, Kranzler H, et al. Genome-wide association study of cannabis dependence severity, novel risk variants, and shared genetic risks. *JAMA Psychiatry* 2016; **73:** 472–80.

5 Pasman JA, Verweij KJH, Gerring Z, et al. GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. *Nat Neurosci* 2018; **21:** 1161–70.

6 Gillespie NA, Neale MC, Jacobson K, Kendler KS. Modeling the genetic and environmental association between peer group deviance and cannabis use in male twins. *Addiction* 2009; **104:** 420–29.

7 Walters RK, Polimanti R, Johnson EC, et al. Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nat Neurosci* 2018; **21:** 1656–69.

8 Kranzler HR, Zhou H, Kember RL, et al. Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. *Nat Commun* 2019; **10:** 1499.

9 Pedersen CB, Bybjerg-Grauholm J, Pedersen MG, et al. The iPSYCH2012 case–cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. *Mol Psychiatry* 2018; **23:** 6–14.

10 Gulcher JR, Kristjánsson K, Gudbjartsson H, Stefánsson K. Protection of privacy by third-party encryption in genetic research in Iceland. *Eur J Hum Genet* 2000; **8:** 739–42.

11 American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th edn. Arlington, VA: American Psychiatric Publishing, 2000.

12 Mors O, Perto GP, Mortensen PB. The Danish psychiatric central research register. *Scand J Public Health* 2011; **39:** 54–57.

13 Lam M, Awasthi S, Watson HJ, et al. RICOPILI: Rapid Imputation for COnsortias PIpeLIne. *Bioinformatics* 2020; **36:** 930–33.

14 Galinsky KJ, Bhatia G, Loh P-R, et al. Fast Principal-Component Analysis reveals convergent evolution of ADH1B in Europe and East Asia. *Am J Hum Genet* 2016; **98:** 456–72.

15 Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; **38:** 904–09.

16 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* 2015; **526:** 68–74.

17 Delaneau O, Marchini J, Zagury J-F. A linear complexity phasing method for thousands of genomes. *Nat Methods* 2011; **9:** 179–81.

18 Hancock DB, Levy JL, Gaddis NC, et al. Assessment of genotype imputation performance using 1000 Genomes in African American studies. *PLoS One* 2012; **7;** e50610.

19 Gudbjartsson DF, Helgason H, Gudjonsson SA, et al. Large-scale whole-genome sequencing of the Icelandic population. *Nat Genet* 2015*;* **47:** 435–44.

20 Demontis, D, Walters RK, Martin J, et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet* 2019; **51:** 63–75.

21 Price AL, Helgason A, Palsson S, et al. The impact of divergence time on the nature of population structure: an example from Iceland. *PLoS Genet* 2009; **5:** e1000505.

22 Bulik-Sullivan BK, Loh P-H, Hilary K Finucane, et al. LD core regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 2015; **47:** 291–95.

23 Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010; **26:** 2190–91.

24 Yang J, Ferreira T, Morris AP, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* 2012; **44:** 369–S3.

25 Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* 2017; **8:** 1826.

26 de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* 2015; **11:** e1004219.

27 Sey NYA, Hu B, Mah W, et al. A computational tool (H-MAGMA) for improved prediction of brain-disorder risk genes by incorporating brain chromatin interaction profiles. *Nat Neurosci* 2020; **23:** 583–93.

28 Lamparter D, Marbach D, Rueedi R, Kutalik Z, Bergmann S. Fast and rigorous computation of gene and pathway scores from SNP-based summary statistics. *PLoS Comput Biol* 2016; **12:** e1004714.

29 Barbeira AN, Dickinson SP, Bonazzola R, et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat Commun* 2018; **9:** 1825.

30 Kreuz DS, Axelrod J. Delta-9-tetrahydrocannabinol: localization in body fat. *Science* 1973; **179:** 391–93.

31 GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 2013; **45:** 580–85.

32 Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet* 2015; **47:** 1236–41.

33 Stinson FS, Ruan WJ, Pickering R, Grant BF. Cannabis use disorders in the USA: prevalence, correlates and co-morbidity. *Psychol Med* 2006; **36:** 1447–60.

34 O'Connor LJ, Price AL. Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nat Genet* 2018; **50:** 1728–34.

35 Zhu Z, Zheng Z, Zhang F, et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat Commun* 2018; **9:** 224.

36 Liu M, Jiang Y, Wedow R, et al. Association studies of up to 1·2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet* 2019; **51:** 237–44.

37 Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; **511:** 421–27.

38 Choi SW, O'Reilly PF. PRSice-2: polygenic risk score software for biobank-scale data. *Gigascience* 2019; **8:** giz082.

39 Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 2005; **102:** 15545–50.

40 Ge T, Chen C-Y, Ni Y, Feng Y-CA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun* 2019; **10:** 1–10.

41 Dennis J, Sealock J, Levinson RT, et al. Genetic risk for major depressive disorder and loneliness in sex-specific associations with coronary artery disease. *Mol Psychiatry* 2019; published online Dec 3. https://doi.org/10.1038/s41380-019-0614-y.

42 Carroll RJ, Bastarache L, Denny JC. R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment. *Bioinformatics* 2014; **30:** 2375–76.

43 Lisdahl, KM, Sher KJ, Conway KP, et al. Adolescent brain cognitive development (ABCD) study: overview of substance use assessment methods. *Dev Cogn Neurosci* 2018; **32:** 80–96.

44 Fischl B. FreeSurfer. *Neuroimage* 2012; **62:** 774–81.

45 Nichols TE. Multiple testing corrections, nonparametric methods, and random field theory. *Neuroimage* 2012; **62:** 811–15.

46 Morris R, Carstairs V. Which deprivation? A comparison of selected deprivation indexes. *J Public Health Med* 1991; **13:** 318–26.

47 Lee JJ, Wedow R, Aysu Okbay A, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1·1 million individuals. *Nat Genet* 2018; **50:** 1112–21.

48 Hancock DB, Guo Y, Reginsson GW, et al. Genome-wide association study across European and African American ancestries identifies a SNP in DNMT3B contributing to nicotine dependence. *Mol Psychiatry* 2018; **23:** 1911–19.

49 Wray NR, Ripke S, Mattheisen M, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet* 2018; **50:** 668–81.

50 Speliotes EK, Willer CJ, Berndt SI, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 2010; **42:** 937–48.

51 Watanabe K, Stringer S, Frei O, et al. A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet* 2019; **51:** 1339–48.

52 Linnér RK, Biroli P, Kong E, et al. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nat Genet* 2019; **51:** 245–57.

53 Lai CSL, Gerrelli D, Monaco AP, Fisher SE, Copp AJ. FOXP2 expression during brain development coincides with adult sites of pathology in a severe speech and language disorder. *Brain* 2003; **126:** 2455–62.

54 Pardiñas AF, Holmans P, Pocklington AJ, et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet* 2018; **50:** 381–89.

55 Fowler IL, Carr VJ, Carter NT, Lewin TJ. Patterns of current and lifetime substance use in schizophrenia. *Schizophr Bull* 1998; **24:** 443–55.

56 McDougle DR, Watson JE, Abdeen AA, et al. Anti-inflammatory ω-3 endocannabinoid epoxides. *Proc Natl Acad Sci USA* 2017; **114:** E6034–43.

57 Snider NT, Nast JA, Tesmer LA, Hollenberg PF. A cytochrome P450-derived epoxygenated metabolite of anandamide is a potent cannabinoid receptor 2-selective agonist. *Mol Pharmacol* 2009; **75:** 965–72.

58 Wagner K, Inceoglu B, Hammock BD. Soluble epoxide hydrolase inhibition, epoxygenated fatty acids and nociception. *Prostaglandins Other Lipid Mediat* 2011; **96:** 76–83.

59 Dudbridge F, Allen RJ, Sheehan NA, et al. Adjustment for index event bias in genome-wide association studies of subsequent events. *Nat Commun* 2019; **10:** 1561.

60 Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. *Genome Med* 2020; **12:** 44.

61 Battistella G, Fornari E, Annoni J-M, et al. Long-term effects of cannabis on brain structure. *Neuropsychopharmacology* 2014; **39:** 2041–48.

62 Mackey S, Allgaier N, Chaaran B, et al. Mega-analysis of gray matter volume in substance dependence: general and substance-specific regional effects. *Am J Psychiatry* 2019; **176:** 119–28.

63 Peterson RE, Kuchenbaecker K, Walters RK, et al. Genome-wide association studies in ancestrally diverse populations: opportunities, methods, pitfalls, and recommendations. *Cell* 2019; **179:** 589–603.

64 Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet* 2019; **51:** 584–91.

65 Smith GD, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. *Int J Epidemiol* 2004; **33:** 30–42.

66 Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am J Epidemiol* 2017; **186:** 1026–34.